Journal of Multivariate Analysis **69**, 10–29 (1999)
Article ID jmva.1998.1788, available online at http://www.idealibrary.com on IDEAL®

Screening among Multivariate Normal Data

Pinyuen Chen

Department of Mathematics, Syracuse University, Syracuse, New York

William L. Melvin

Georgia Tech Research Institute, Smyrna, Georgia

and

Michael C. Wicks

United States Air Force Research Laboratory, Rome, New York

This paper considers the problem of screening k multivariate normal populations (secondary data) with respect to a control population (primary data) in terms of covariance structure. A screening procedure, developed based upon statistical ranking and selection theory, is designed to include in the selected subset those populations which have the same (or similar) covariance structure as the control population, and exclude those populations which differ significantly. Formulas for computing the probability of a correct selection and the least favorable configuration are developed. The sample size required to achieve a specific probability requirement is also developed, with results presented in tabular form. This secondary data selection procedure is illustrated via an example with applications to radar signal processing. © 1999 Academic Press

AMS 1991 subject classifications: 62E15, 62F07, 62H10.

Key words and phrases: hypergeometric function in matrix argument, indifference zone approach, eigenvalue, least favorable configuration, multivariate normal, probability of a correct screening, radar signal processing, ranking and selection, subset selection approach.

1. INTRODUCTION

Ranking and selection procedures are generally developed using either an indifference zone or a subset selection approach. The literature on ranking and selection theory is dominated by these two methods. In this paper, a variation of the subset selection approach is used to develop a screening procedure for choosing secondary data. By partitioning secondary data into two groups based upon a quantitative measure of similarity in covariance structure, samples which differ significantly from a control population can be discarded. The remaining samples are then

0047-259X/99 $30.00

processed via standard statistical hypothesis testing techniques to determine the presence or absence of targets. This result is a new approach to radar signal processing with dramatically improved performance over conventional techniques.

Consider k independent populations $\pi_1, \pi_2, ..., \pi_k$ where the underlying distribution of π_i is F_{θ_i}, $i=1, 2, ..., k$. The unknown real parameter, θ_i, $i=1, 2, ..., k$, represents the value of a quantity of interest for the ith population. By definition, we select population π_i over π_j if θ_i is greater than θ_j. The ordered values of θ_i for all i are denoted by $\theta_{[1]} \leqslant \cdots \leqslant \theta_{[k]}$. In general, $\theta_{[i]} \neq \theta_i$. One approach to solving the basic problem of selecting the *best* population, called the indifference zone formulation, was developed in Bechhofer (1954). In Bechhofer's paper, the selection of the population associated with the ranked parameter $\theta_{[k]}$ results in a correct selection (CS). For the indifference zone approach to be of value, the procedure R must establish a lower bound on the probability of a correct selection $P(\mathrm{CS})$. The minimum value of $P(\mathrm{CS})$ is P^*, with $1/k < P^* < 1$ whenever the separation between $\theta_{[k]}$ and $\theta_{[k-1]}$ exceeds some minimum specified value. Let $\delta(\theta_i, \theta_j)$ denote an appropriate nonnegative measure of the separation between the population associated with θ_i and θ_j. For the minimum probability of correct selection, P^*, δ^* is the minimum separation distance. For any specified $\delta^* > 0$, let the preference zone, Ω_{δ^*} be the subset of the parameter space $\Omega = \{\tilde{\theta} \mid \tilde{\theta} = (\theta_1, ..., \theta_k)\}$ defined by $\Omega_{\delta^*} = \{\tilde{\theta} \mid \delta(\theta_{[k]}, \theta_{[k-1]}) \geqslant \delta^*\}$. Let $P(\mathrm{CS} \mid \mathrm{R})$ denote the probability of a correct selection under the procedure R. In order for this procedure to be valid, it should satisfy $P(\mathrm{CS} \mid \mathrm{R}) \geqslant P^*$ for all $\tilde{\theta} \in \Omega_{\delta^*}$. The complement of the preference zone Ω_{δ^*} is called the indifference zone, a subset of the parameter space where no requirement on $P(\mathrm{CS})$ is made. For the analysis of measured data, δ^* and P^* are specified in advance. Suppose that the procedure R is based on samples of fixed size n from each population. One problem of practical interest in radar signal processing is to determine the smallest sample size n for which the probability requirement P^* holds.

In the subset selection approach of Gupta (1956), a procedure was developed to guarantee a non-empty subset of the k given populations which include the desired (or best) population with a minimum probability P^*. Any subset which includes the desired population results in a correct selection. In case of a tie, any contender may be tagged best. Any valid procedure R should satisfy $P(\mathrm{CS} \mid \mathrm{R}) \geqslant P^*$ for all $\tilde{\theta} \in \Omega$. In the subset selection approach, the size of the selected subset S is not decided in advance, but is determined based on the analysis of data.

The procedures developed in ranking and selection theory are designed to satisfy the requirement for a minimum probability of a correct selection P^*. Any parameter configuration $\tilde{\theta}$ which yields the infimum of the $P(\mathrm{CS})$ over Ω_{δ^*} in the indifference zone approach, or Ω in the subset selection approach, is called the least favorable configuration (LFC). Many variations and generalizations of these two basic approaches have been studied. For example, one problem involves procedures for selecting the most appropriate sample populations better than a control population π_0. These sample populations may then be used to estimate other parameters of interest such as the covariance matrix. In our study of selection procedures for analyzing radar data, the control population can be taken as the primary data (under the null hypothesis). The secondary data are selected from those independent populations $\pi_1, \pi_2, \ldots, \pi_k$ having the same or similar covariance structure as the control population. This approach forms the basis for a solution to the nonhomogeneous interference problem in radar signal processing.

Radar data in the form of random vectors or matrices are processed statistically via the techniques developed in multivariate analysis. In Section 2, the application of ranking and selection theory to radar signal processing is discussed. In Section 3, a screening procedure and method to solve the non-homogenous problem in multivariate analysis is presented. A derivation of the probability of a correct selection P(CS) for this screening procedure, and its least favorable configuration, are included in Section 4. Numerical results and sample size requirements are provided in Section 5. An example with radar signal processing applications is also included. Final remarks and future research are discussed in Section 6.

2. APPLICATIONS OF RANKING AND SELECTION THEORY TO RADAR SIGNAL PROCESSING

Statistical inferences in radar signal processing involves two of the major focus areas in statistical research, hypothesis testing and multivariate analysis. The null hypothesis, mean vector equal to zero (i.e., $H_0 : \tilde{\mu} = 0$) implies target absent in the test cell. The alternate hypothesis, nonzero mean vector (i.e., $H_0 : \tilde{\mu} = \xi$), implies target present in the test cell. For example, Khatri and Rao (1987) developed a test of the mean vector given that only an estimate of the covariance matrix is available. Since measured radar data are assumed to be correlated complex-valued vectors, multivariate analysis is appropriate for statistical hypothesis testing. Although recent research in radar signal processing focus on a wide variety of density functions, the focus of this paper is on multivariate normal theory.

Classical detection theory was developed under the Gaussian assumption. As such, target returns are embedded in homogeneous Gaussian interference. The term "homogeneous" refers to the covariance matrices of the reference cells (the secondary data) which are assumed to have the same structure as the covariance matrix of the test cell (the primary data) under the null hypothesis. Under this assumptions, the likelihood ratio tests has been well studied and documented (see for examples, Kelly (1986) and Khatri and Rao (1987)). As it has been reported in Melvin *et al.* (1998), the detection of targets in heterogeneous (nonhomogenous) interference is an important and challenging signal processing research problem. Classical detection processing techniques, developed under the homogeneous assumption, may suffer a significant loss in performances if the true interference environment is heterogeneous.

Recently, Raghavan *et al.* (1995) addressed several interesting aspects of target detection in the present of nonhomogenous interference. In their analysis, the correlation properties of the interference remains unknown. However, distribution theories needed for computing measures of performance (probabilities of false alarm and detection) for their proposed test statistic are incomplete. The probability in the null hypothesis (probability of false alarm) is computed assuming the target signal is deterministic, thus causing no change in second order statistic (covariance matrix). The probability in the alternate hypothesis (probability of detection) was not computed for the proposed test statistic. Instead, an asymptotic performance analysis, applicable only under the large sample assumption was conducted. Strictly speaking, a complete solution to this problem paralleling the development of the Generalized Likelihood Ratio (GLR) test of Kelly (1986) should include the distribution theory for the test statistic $z^H S^{-1} z$, while z is normal (and zero mean under the null hypothesis) and nS, where n is the sample size, is a Wishart, independent of z. Then, a comparison between this test and the GLR test would be possible.

An alternative approach is to determine whether the covariance structure of each reference cell (secondary data) is the same as the control data (test cell) under the null hypothesis. If a test of homogeneity is the desired result, further analysis is not needed. However, in radar signal processing, as with many other applications, we are not only interested in which secondary data have the same covariance structure as that of the test cell. We are also interested in selecting those reference cells whose covariance structure is the same as that of the test cell, or similar, that traditional algorithms for detection processing may be applied. Next, we propose a screening procedure to solve this problem.

A procedure for eliminating secondary data which differs significantly in covariance structure from the primary data is needed. Suppose that the secondary data is obtained from the independent random vectors $Y_1, Y_2, ..., Y_K$ and the primary data from the random vector Y_0. Assume also that we can segregate the vectors $Y_1, Y_2, ..., Y_K$ into k subgroups $\{Y_{11}, ..., Y_{1m}\}, \{Y_{21}, ..., Y_{2n}\}, ..., \{Y_{k1}, ..., Y_{kp}\}$, where the vectors from each subgroup are from the same (or similar) populations. This is true for the analysis of radar data collected during controlled flights, since environmental features (lakes, rivers, forest, etc.) and cultural features (highways, urban developments, etc.) are well documented. The goal is to eliminate those subgroups with a covariance structure which is significantly different from that of the primary data. To achieve this goal, we need to modify the classical subset selection approach in such a way that the concept of comparing with a control is included in the basic formulation of the test. The new formulation, which will be called screening with respect to a control, will be formally defined in the next section. Under this formulation we will propose a procedure to select the similar populations and to eliminate the dissimilar populations simultaneously.

Selection among multivariate normal populations was thoroughly reviewed by Gupta and Panchapakesan (1979) for indifference zone formulation in Chapt. 7, for subset selection formulation in Chapt. 14, and for comparison with a control in Sect. 20.8. The formulation that is the closest to our screening formulation, to be defined in next section, is by Krishnaiah (1967), who considered the problem of selecting multivariate normal populations better than a control on the basis of linear combinations of the elements of covariance matrices. The formula of the probability of a correct screening $P(\mathrm{CS})$ for our procedure, to be dereived in Section 4, has a form similar to but more general than the $P(\mathrm{CS})$ of Gupta and Panchapakesan (1969) who considered selecting a subset containing the population associated with the largest multiple correlation coefficients.

3. SCREENING WITH RESPECT TO A CONTROL

We begin with some notations and basic definitions. If $A=(a_{ij})$ is a matrix of complex numbers, the conjugate transpose of A is defined by $A^*=(a_{ji}^*)$, where a_{ji}^* is is the complex conjugate of a_{ij}. A square complex matrix is said to be Hermitian if $A=A^*$. Let A be an $n\times n$ Hermitian matrix and let x be any $n\times 1$ vector. Then x^*Ax is called a Hermitian form of A, and it will be a real number. A will be said to be positive definite iff $x^*Ax>0$ for all nonzero vectors x. If A is positive definite, then all the eigenvalues are positive. (See, for example, Theorem 1.9.1 in Srivastava and Khatri (1979).) Let $z=x+iy$ be a complex random p vector with mean θ and covariance matrix $Q=\Sigma_1+i\Sigma_2$. Then z has a complex multivariate normal distribution with mean θ and covariance matrix Q, written as $z\sim \mathrm{CN}_p(\theta, Q)$ iff $\binom{x}{y}\sim N_{2p}(\binom{\theta_1}{\theta_2}, \Sigma)$ where $\Sigma=\frac{1}{2}\begin{pmatrix}\Sigma_1 & -\Sigma_2\\ \Sigma_2 & \Sigma_1\end{pmatrix}$, $\theta=\theta_1+i\theta_2$. (See, for example, Definition 2.9.2 in Srivastava and Khatri (1979).)

Let $\pi_1, \pi_2, \ldots, \pi_k$ represent k multivariate (p-variate) complex normal populations $\mathrm{CN}_p(\mu_i, \Sigma_i)$, $i=1, 2, \ldots, k$, and let π_0 be a control p-variate complex normal population $\mathrm{CN}_p(\mu_0, \Sigma_0)$. Assume that $\mu_i=0$, $i=0, 1, 2, \ldots, k$, since our primary concern is with the structure of the covariance matrix. In radar signal processing, there are usually several guard cells near the test cell or primary data. Thus, it is possible to use one or more guard cells as a control population. For the comparison of two univariate normal populations of zero mean, a measure of similarity is the ratio function, $d(x, y) = x/y$, since the variance is a scale parameter. The covariance matrix of a multivariate normal random vector has similar properties as the variance of univariate normal random variable, especially in distribution theory. Thus we will also use the ratio of two covariance matrices as a distance measure in our study. Let $\lambda_{i,1} \geqslant \lambda_{i,2} \geqslant \cdots \geqslant \lambda_{i,p} > 0$ denote the ordered eigenvalues of $\Sigma_0 \Sigma_i^{-1}$. Now we define the two disjoint and exhaustive subsets Ω_G and Ω_B of the set $\Omega = \{\pi_1, \pi_2, ..., \pi_k\}$, by using a pair of distance functions d_1 and d_2 defined as

$$d_1(\Sigma_i, \Sigma_0) = \lambda_{i,1} \qquad \text{and} \qquad d_2(\Sigma_i, \Sigma_0) = \lambda_{i,p}, \tag{3.1}$$

and

$$\Omega_B = \{\pi_i \mid d_1(\Sigma_i, \Sigma_0) < \delta_1^* \quad \text{or} \quad d_2(\Sigma_i, \Sigma_0) > \delta_2^*\} \quad \text{and} \quad \Omega_G = \Omega - \Omega_B, \tag{3.2}$$

where $\delta_1^* < \delta_2^*$ are preassigned positive real numbers used to differentiate between similar and dissimilar populations. Theoretically, the values of δ_1^* should be less than 1 and the value of δ_2^* should be greater than 1 since $\delta_1^* = \delta_2^* = 1$ is equivalent to the perfect case when the control population has exactly the same covariance matrix as that of the experimental populations. A population is considered similar to a control population where the distance measures approaching unity. Our goal is to separate the populations obtained from the reference data into two disjoint subsets, S_G and S_B. The separation is correct if $S_G \subset \Omega_G$, meaning that all populations included in selected subset S_G have similar covariance structure as the control population. It also means that all populations with significantly different covariance structures are eliminated. We require a procedure R that will satisfy a predetermined probability requirement $P(\mathrm{CS} \mid \mathrm{R}) \geqslant P^*$. The proposed procedure, R_c, is defined as:

Procedure R_c. For each population Π_i $(i = 1, 2, ..., k)$, computer $T_i = (x^H S_i^{-1} x)/n$ where n is the common sample size, x is the primary data vector, and S_i is the sample covariance matrix associated with population Π_i. Partition the set of populations $\Omega = \{\pi_1, \pi_2, ..., \pi_k\}$ into two subsets S_G and S_B. The subset S_G consists of those populations Π_i with $c \leqslant T_i \leqslant d$ where c and d are chosen such that the probability requirement $P(\mathrm{CS}) \geqslant P^*$ is satisfied.

The selection statistic in procedure R_c, $T_i = (x^H S_i^{-1} x)/n$, seems to be a reasonable choice for estimating scalar functions of the matrix $\Sigma_0 \Sigma_i^{-1}$ when only one observation is available in the control population Π_0. Considering S_i^{-1} is the inverse of the covariance matrix, then the measure used in the selection procedure could be thought of as a ratio in matrix form.

4. THE PROBABILITY OF A CORRECT SCREENING FOR THE PROPOSED PROCEDURE R_C

In ranking and selection theory, we use $P(\mathrm{CS})$ to measure the performance of a selection procedure. In this analysis, our goal is to screen populations according to their covariance structure. A natural measure of the performance of the proposed screening procedure is $P(\mathrm{CS})$. Next, we derive the distribution of the selection statistic $T_i = (x^H S_i^{-1} x)/n$ in procedure R_c and the least favorable configuration (LFC). Also computed is the minimum of $P(\mathrm{CS})$ under the LFC, in terms of the multivariate normal density function and the chi-squared, density function. In order to implement this procedure, a numerical solution to the integral equation, $\min\{P(\mathrm{CS} \mid \mathrm{LFC})\} = P^*$, is required.

4.1. *The Distribution of* $T_i = (x^H S_i^{-1} x)/n$

Assume that $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_k$ are k independent random vectors from a p-variate complex normal distribution $\mathrm{CN}_p(0, R_1)$. The random vector $\mathbf{X}_{k+1}$ is distributed, independently of $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_k$, according to a p-variate complex normal distribution $\mathrm{CN}_p(0, R_2)$. Both R_1 and R_2 are unknown.

Our selection statistic can be written as $T = (X_{k+1}^H S_1^{-1} X_{k+1})/n$ where $\mathbf{S}_1$ is the sample variance–covariance matrix of $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_k\}$. The selection statistic can be written

$$
\begin{aligned}
nT &= \mathbf{X}_{k+1}^H \mathbf{S}_1^{-1} \mathbf{X}_{k+1} = \mathrm{tr}(\mathbf{X}_{k+1}^H \mathbf{S}_1^{-1} \mathbf{X}_{k+1}) \\
&= \mathrm{tr}(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^H \mathbf{S}_1^{-1}) = \mathrm{tr}(\mathbf{S}_2 \mathbf{S}_1^{-1}), \qquad \text{where} \quad \mathbf{S}_2 = \mathbf{X}_{k+1} \mathbf{X}_{k+1}^H \qquad (4.1)
\end{aligned}
$$

First, consider real normal random variables only. Since S_2 has rank 1 and S_1 has full rank p, the rank of $\mathbf{S}_2 \mathbf{S}_1^{-1}$ is 1. Thus

$$
nT = \mathrm{tr}(\mathbf{S}_2 \mathbf{S}_1^{-1}) = \text{the largest eigenvalue of } \mathbf{S}_2 \mathbf{S}_1^{-1} \qquad (4.2)
$$

For the non-singular case where both S_2 and S_1 have full rank p and orders n_1 and n_2, respectively, formula (16) of Khatri (1967) gives the density of f_1, the largest eigenvalue of $\mathbf{S}_2\mathbf{S}_1^{-1}$, as

$$c\,|\Lambda|^{-(1/2)\,n_1} f_1^{(1/2)\,pn_1-1}\,|I+f_1\Lambda^{-1}|^{-(1/2)(n_1+n_2)}$$
$$\times\, 3F_2(\tfrac{1}{2}n_1+\tfrac{1}{2}n_2,\tfrac{1}{2}p+1,\tfrac{1}{2}p-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}(n_1+p+1);\, f_1(\Lambda+f_pI)^{-1}), \quad (4.3)$$

where

$$c=\Gamma(\tfrac{1}{2})\,\Gamma_p(\tfrac{1}{2}\mathbf{n}_1+\tfrac{1}{2}\mathbf{n}_2)\,\Gamma_{p-1}(\tfrac{1}{2}\mathbf{p}+1)\{\Gamma(\tfrac{1}{2}\mathbf{p})$$
$$\times\,\Gamma(\tfrac{1}{2}\mathbf{n}_1)\,\Gamma_p(\tfrac{1}{2}\mathbf{n}_2)\,\Gamma_{p-1}(\tfrac{1}{2}\mathbf{n}_1+\tfrac{1}{2}\mathbf{p}+\tfrac{1}{2})\}^{-1}$$

$$\Gamma_p(\mathbf{t})=\pi^{(1/4)\,p(p-1)}\prod_{j=1}^{p}\Gamma(\mathbf{t}-\tfrac{1}{2}\mathbf{j}+\tfrac{1}{2}), \qquad \Lambda=\mathbf{diag}(\lambda_1,\lambda_2,\ldots,\lambda_p),$$

where λ s are the eigenvalues of $\mathbf{R}_2\mathbf{R}_1^{-1}$, and ${}_3\mathbf{F}_2$ is the hypergeometric function in matrix argument as defined in James (1964). Since the derivation of the distribution of T involves the simplification of ${}_3\mathbf{F}_2$, the complete definition of the hypergeometric function in matrix argument is required. Consider the partition of $\mathbf{k}=(\mathbf{k}_1,\mathbf{k}_2,\ldots,\mathbf{k}_m)$, $\mathbf{k}_1\geqslant\mathbf{k}_2\geqslant\cdots\geqslant\mathbf{k}_m\geqslant 0$, $\mathbf{k}_1+\mathbf{k}_2+\cdots+\mathbf{k}_m=\mathbf{k}$, into at most m (the number of variables in each vector) parts. Then from p. 477 of James (1964), we have, by definition,

$${}_p\mathbf{F}_q(\mathbf{a}_1,\ldots,\mathbf{a}_p;\mathbf{b}_1,\ldots,\mathbf{b}_q;\mathbf{S})=\sum_{k=0}^{\infty}\sum_{\mathbf{k}}\frac{(\mathbf{a}_1)_\mathbf{k}\cdots(\mathbf{a}_p)_\mathbf{k}}{(\mathbf{b}_1)_\mathbf{k}\cdots(\mathbf{b}_q)_\mathbf{k}}\frac{\mathbf{C}_\mathbf{k}(\mathbf{S})}{\mathbf{k}!}, \quad (4.4)$$

where $\mathbf{a}_1,\ldots,\mathbf{a}_p,\mathbf{b}_1,\ldots,\mathbf{b}_q$ are real or complex constants and the multivariate hypergeometric coefficient

$$(\mathbf{a})_\mathbf{k}=\prod_{i=1}^{m}(\mathbf{a}-\tfrac{1}{2}(\mathbf{i}-1))_{k_i},$$
$$\text{where}\quad (\mathbf{a})_k=\mathbf{a}(\mathbf{a}+1)(\mathbf{a}+2)\cdots(\mathbf{a}+\mathbf{k}-1). \quad (4.5)$$

It is clear from the arguments made in Section 13.2.4 of Anderson (1984), that in order to obtain the density function of the roots of $|A-f(A+B)|$ for the singular case (that is, when the rank of A is less than p), we need only to change the density function of the roots of $|\boldsymbol{A}-\mathbf{f}(\boldsymbol{A}+\mathbf{B})|$ by changing variables in formula (4.3), replacing (n_1,n_2,p) to (p,n_1,n_1+n_2-p), to obtain

$$c_p\,|\Lambda|^{-(1/2)\,p} f_1^{(1/2)\,pn_1-1}\,|I+f_1\Lambda^{-1}|^{-(1/2)(n_1+n_2)}$$
$${}_3F_2(\tfrac{1}{2}n_1+\tfrac{1}{2}n_2,\tfrac{1}{2}n_1+1,\,\tfrac{1}{2}n_1-\tfrac{1}{2};\,\tfrac{1}{2}n_1,\,\tfrac{1}{2}(n_1+p+1);\,f_1(\Lambda+f_1I)^{-1}), \quad (4.6)$$

where

$$\begin{aligned} c_p &= \Gamma(\tfrac{1}{2})\, \Gamma_{n_1}(\tfrac{1}{2}n_1+\tfrac{1}{2}n_2)\, \Gamma_{n_1-1}(\tfrac{1}{2}n_1+1)\{\Gamma(\tfrac{1}{2}n_1) \\ &\quad \times \Gamma(\tfrac{1}{2}p)\, \Gamma_{n_1}(\tfrac{1}{2}(n_1+n_2-p)\, \Gamma_{n_1-1}(\tfrac{1}{2}n_1+\tfrac{1}{2}p+\tfrac{1}{2})\}^{-1} \\ &= \Gamma(\tfrac{1}{2})\, \Gamma_{n_1}(\tfrac{1}{2}n_1+\tfrac{1}{2}n_2)\{\Gamma(\tfrac{1}{2}n_1)\, \Gamma(\tfrac{1}{2}p) \\ &\quad \times \Gamma_{n_1}(\tfrac{1}{2}(n_1+n_2-p)\, \Gamma_{n_1-1}(\tfrac{1}{2}n_1+\tfrac{1}{2}p+\tfrac{1}{2})\}^{-1} \end{aligned}$$

and Λ is the diagonal matrix with n_1 elements $\lambda_1 > \lambda_2 > \cdots > \lambda_{n_1}$.

Now replace n_1 by 1 in ${}_3\mathbf{F}_2$ in (4.6). The third argument in the first 3-vector becomes 0.

For any $k>0$, since $k_1>0$, we have

$$(\mathbf{a}_3)_{\mathbf{k}} = (0)_{\mathbf{k}} = (0)_{k_1}(-\tfrac{1}{2})_{k_2} \cdots = 0. \tag{4.7}$$

For $k=0$, by Formula (4) in James (1964), we get

$$(a)_k = 1 \qquad \text{for all } a. \tag{4.8}$$

Thus we obtain

$$\begin{aligned} &{}_3F_2(\tfrac{1}{2}n_1+\tfrac{1}{2}n_2, \tfrac{1}{2}n_1+1, \tfrac{1}{2}n_1-\tfrac{1}{2}; \tfrac{1}{2}n_1, \tfrac{1}{2}(n_1+p+1); f_1(\Lambda+f_1 I)^{-1}) \\ &\quad = {}_3F_2(\tfrac{1}{2}+\tfrac{1}{2}n_2, \tfrac{3}{2}, 0; \tfrac{1}{2}, \tfrac{1}{2}(p+2); f_1(\Lambda+f_1 I)^{-1}) \\ &\quad = \sum_{\mathbf{k}} C_{\mathbf{k}}(f_1(\Lambda+f_1 I)^{-1}), \end{aligned} \tag{4.9}$$

where $\mathbf{k}$ is the partition for $k=0$. Moreover by (17) in James (1964),

$$(\operatorname{tr} S)^k = \sum_{\mathbf{k}} C_{\mathbf{k}}(S). \tag{4.10}$$

Thus ${}_3F_2$ in (4.9) is 1. Now by letting $n_1=1$ in the density function (4.6), we finally obtain the density of $T=f_1$:

$$f(f_1) = c_p\, |\Lambda|^{-p/2}\, f_1^{(p-2)/2}\, |I+f_1\Lambda^{-1}|^{-(n_1+n_2)/2}. \tag{4.11}$$

Considering the complex case via heuristic arguments, the analogous theory for the density function of the largest eigenvalue of the matrix $S_2 S_1^{-1}$ does not exist. However, based on (1) the fact that both S_2 and S_1 are positive Hermitian with positive eigenvalues, as are the eigenvalues of $S_2 S_1^{-1}$, (2) the F distribution can be derived from the ratio of two Chi-squared random variables; and (3) most of the Chi-square random variables involved in the complex case have twice as many degrees of freedom as in the real case, we obtain from p. 188 of Khatri and Rao

(1987) that the density function for the complex case has the following density:

$$f(f_1) = c_p^* |\Lambda|^{-p} f_1^{p-2} |I + f_1 \Lambda^{-1}|^{-(n_1+n_2)}$$
$$\text{with appropriate coefficient} \quad c_p^*. \tag{4.12}$$

Let $\mathbf{X}_1(p \times n_1)$ and $\mathbf{X}_2(p \times n_2)$ be independent matrix variates, columns of $\mathbf{X}_1$ being independently distributed as $N(0, R_1)$ and those of $\mathbf{X}_2$ independently distributed as $N(0, R_2)$. In testing the equality of R_1 and R_2, various criteria have appeared in the literature. Let $f_1 \geqslant f_2 \geqslant \cdots \geqslant f_p$ be the eigenvalues of $|\mathbf{X}_1\mathbf{X}_1' - f\mathbf{X}_2\mathbf{X}_2'| = 0$ and let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$ be the eigenvalues of $|R_1 - fR_2| = 0$. Roy (1945) considered f_1. Hotelling (1951) considered $U = \sum_{i=1}^p f_i$. As for our problem, $n_1 = 1$. Thus $f_2 = f_3 = \cdots = f_p = 0 < f_1$. Our selection statistic T is exactly the same as Roy's and Hotelling's test statistic.

4.2. *P*(*CS*) *and LFC*

Assume that there are $k_1 + k_2$ bad populations. Among them, k_1 populations $\Pi_1, \Pi_2, ..., \Pi_{k_1}$ are *significantly better* than the control and k_2 populations $\Pi_{k-k_2}+1, \Pi_{k-k_2}+2, ..., \Pi_k$ are *significantly worse* than the control. That is, for $i = 1, 2, ..., k_1$, $d_1(\Sigma_i, \Sigma_0) = \lambda_{i,1} < \delta_1^*$. For $i = k - k_2 + 1, ..., k$, $d_2(\Sigma_i, \Sigma_0) = \lambda_{i,p} > \delta_2^*$. As a consequence, there are $k - k_1 - k_2$ similar populations. We also assume that each population has sample size n. It can be seen from the density function, given in formula (4.12), of our test statistic $T_i = nx^H S_i^{-1} x$ that the distribution depends on Σ_0 and Σ_1 only through $\Lambda = \sigma_0 \Sigma_1^{-1}$. Thus in the following derivation of probability of a correct screening, we will assume that $\Sigma_0 = I$, and for each $i = 1, 2, ..., k_1 + k_2$, we assume that the population covariance matrix $\Sigma_{1,i} = \text{diag}(1/\lambda_{i1}, 1/\lambda_{i2}, ..., 1/\lambda_{ip})$. Then we have

$$\begin{aligned}
P(\text{CS}) &= P\ (\text{all the dissimilar populations are eliminated}) \\
&= P(T_i \notin [c, d] \text{ for } i = 1, ..., k_1 \text{ and } i = k - k_2 + 1, ..., k) \\
&\geqslant P(T_i < c, i = 1, 2, ..., k_1;\ T_j > d, j = k - k_2 + 1, ..., k) \\
&= \int_x P((x^H S_i^{-1} x)/n < c, i = 1, 2, ..., k_1; \\
&\qquad (x^H S_j^{-1} x)/n > d, j = k - k_2 + 1, ..., k)\ \varphi(x)\, dx
\end{aligned} \tag{4.13}$$

where x is a p-dimension complex normal random vector with 0 mean vector and covariance matrix I and $\varphi(x)$ is its density function. From Rao (1973, p. 538), for given x, $x^H \Lambda_i x / [(x^H S_i^{-1} x)/n]$ follows a chi-squared distribution with $2(n-p+1)$ degrees of freedom. Thus since Λ is positive definite and $x^H \Lambda x > 0$ for all x, we obtain

$$\begin{aligned} P((x^H S_i^{-1} x)/n < c,\ i=1, 2, ..., k_1;\ (x^H S_j^{-1} x)/n > d,\ j=k-k_2+1, ..., k) \\ = P(x^H \Lambda_i x/[(x^H S_i^{-1} x)/n] > x^H \Lambda_i x/c,\ i=1, 2, ..., k_1; \\ x^H \Lambda_j x/[(x^H S_j^{-1} x)/n] < x^H \Lambda_j x/d,\ j=k-k_2+1, ..., k) \end{aligned} \tag{4.14}$$

If we denote $x^H \Lambda_i x/[(x^H S_i^{-1} x)/n]$ by Y_i for all $i=1, 2, ..., k_1+k_2$, we obtain from (4.14) that

$$\begin{aligned} P((x^H S_i^{-1} x)/n < c,\ i=1, 2, ..., k_1;\ (x^H S_j^{-1} x)/n > d,\ j=k-k_2+1, ..., k) \\ = P(Y_i > x^H \Lambda_i x/c,\ i=1, 2, ..., k_1; \\ Y_j < x^H \Lambda_j x/d,\ j=k-k_2+1, ..., k). \end{aligned} \tag{4.15}$$

Since the value of $x^H \Lambda_i x$ increases as we increase any diagonal element of Λ_i when all the other diagonal elements are held fixed, the probability in (4.15) reaches its minimum when all the diagonal elements in Λ_i for $i=1, 2, ..., k_1$ are increased to their maximum possible value δ_1^* and all the diagonal elements in Λ_j for $j=k-k_2+1, ..., k$ are decreased to their minimum possible value δ_2^*. It is also clear that the minimum of $P(\mathrm{CS})$ occurs when the total number of dissimilar populations equals k. Therefore we have complete the proof of the following theorem.

THEOREM 4.1. *The least favorable configuration for any parameter vector* $(\lambda_{1,1}, \lambda_{1,2}, ..., \lambda_{1,p}; \lambda_{2,1}, \lambda_{2,2}, ..., \lambda_{2,p}; ...; \lambda_{k,1}, \lambda_{k,2}, ..., \lambda_{k,p})$ *under procedure* R_C *is given by*

$$\begin{aligned} \lambda_{i,1} = \lambda_{i,2} = \cdots = \lambda_{i,p} = \delta_1^*, \qquad i=1, 2, ..., m; \\ \lambda_{j,1} = \lambda_{j,2} = \cdots = \lambda_{j,p} = \delta_2^*, \qquad j=m+1, ..., k, \end{aligned} \tag{4.16}$$

where m *is an integer between* 0 *and* k *that minimizes the probability of a correct selection in Eq.* (4.13).

As a consequence of the above theorem, under the LFC in (4.16), the sample covariance matrix S_i for $i=1, 2, ..., m$ follows a complex Wishart distribution with parameters n and I/δ_1^* and the sample covariance matrix S_j for $j=m+1, ..., k$ follows a complex Wishart distribution with parameters n and I/δ_2^*. Furthermore the test statistic Ts, under the least favorable configuration, follow the following distribution:

$$\left(\frac{n-p+1}{p}\right)\frac{T_i}{\delta_1^*}, \qquad i=1, 2, ..., m \text{ follows an F distribution}$$
$$\text{with } 2p \text{ and } 2(n-p+1) \text{ degrees of freedom} \tag{4.17}$$

and

$$\left(\frac{n-p+1}{p}\right)\frac{T_j}{\delta_2^*}, \qquad j=m+1, m+2, ..., k \text{ follows an F distribution}$$
$$\text{with } 2p \text{ and } 2(n-p+1) \text{ degrees of freedom.}$$

From the above theorem, we can write the minimum of $P(\mathrm{CS})$ as

$$\min\{P(\mathrm{CP})\} = P(T_i < c, i=1, 2, ..., m; T_j > d, j=m+1, ..., k)$$
$$\text{where } T\text{'s satisfy (4.17).} \tag{4.18}$$

In order to implement the proposed procedure R_C, procedure parameters c and d as a function of n, p, k, δ_1^*, and δ_2^* are required. Although Eq. (4.18) provides a simple expression for $\min\{P(\mathrm{CS})\}$, in order to avoid computations involving correlated F distributions, we will use conditional arguments to rewrite $\min\{P(\mathrm{CS})\}$ in terms of multivariate normal function and chi-squared density functions.

We obtain the same minimum $P(\mathrm{CS})$ as expressed in Eq. (4.18) if the control population has covariance matrix $\Sigma_0 = I$ and the population covariance matrices are

$$\Sigma_i = \begin{bmatrix} 1/\delta_1^* & 0 & \cdot & \cdot & \cdot \\ 0 & 1/\delta_1^* & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 1|\delta_1^* \end{bmatrix}.$$

for $i=1, 2, ..., m$; and

$$\Sigma_i = \begin{bmatrix} 1/\delta_2^* & 0 & \cdot & \cdot & \cdot \\ 0 & 1/\delta_2^* & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 1|\delta_2^* \end{bmatrix}.$$

for $i=m+1, m+2, \ldots, k$. Therefore, from (4.18) and Theorem 4.1,

$$P(\mathrm{SC} \mid \mathrm{LFC}) = P((x^H S_i^{-1} x)/n < c,\ i=1, 2, \ldots, m;$$
$$(x^H S_j^{-1} x)/n > d,\ j=m+1, \ldots, k), \tag{4.19}$$

where x is a p-variate complex normal random vector CN(0, 1), S_i is the sample covariance matrix with covariance matrix Σ_i, and S_j is the sample covariance matrix with covariance matrix Σ_j. Define $x=(x_1, x_2, \ldots, x_p)$, where the x_i's are independent and identically distributed (iid) univariate complex CN(0, 1). From Rao (1973, p. 538), the statistic $W_i = L^H \Sigma_i^{-1} L / [(L^H S_i^{-1} L)/n]$ follows a chi-squared distribution with $2(n-p+1)$ degrees of freedom for any fixed vector L. Thus the statistic $L^H S_i^{-1} L/n = (L^H \Sigma_i^{-1} L)/W_i$ is distributed as the reciprocal of a chi-squared random variable times a constant. From Eq. (4.19),

$$P(\mathrm{CS} \mid \mathrm{LFC}) = P\{(x^H \Sigma_i^{-1} x/[(x^H S_i^{-1} x)/n)] > (x^H \Sigma^{-1}{}_i x)/c,\ i=1, 2, \ldots, m;$$
$$(x^H \Sigma_j^{-1} x/[(x^H S_j^{-1} x)/n] < (x^H \Sigma_j^{-1} x)/d,$$
$$j=m+1, m+2, \ldots, k\}$$
$$= \iint \cdots \int P\{W_i > (\delta_1^*/c)[x_1\tilde{x}_1 + x_2\tilde{x}_2 + \cdots + x_p\tilde{x}_p],$$
$$i=1, 2, \ldots, m;$$
$$W_j < (\delta_2^*/d)[x_1\tilde{x}_1 + x_2\tilde{x}_2 + \cdots + x_p\tilde{x}_p],$$
$$j=m+1, \ldots, k\}\ \phi(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \cdots dx_p, \tag{4.20}$$

where W_i, $i=1, \ldots, m$ and W_j, $j=m+1, \ldots, k$ are iid chi-squared random variables with $2(n-p-1)$ degrees of freedom for a given x, $\tilde{x}_i$, $i=1, 2, \ldots, p$ is the conjugates of x_i, and $\phi(x_1, x_2, \ldots, x_p)$ is the probability density function of a p-varaiate CN(0, I). This probability depends on $x_1, x_2, \ldots, x_p$ only through $x_1\tilde{x}_1 + x_2\tilde{x}_2 + \cdots + x_p\tilde{x}_p$ which has a chi-squared distribution with $2p$ degrees of freedom. Thus

$$P(\mathrm{CS} \mid \mathrm{LFC}) = \int_0^\infty (\chi^2(\delta_2^* y/d))^{k-m}\, (1-\chi^2(\delta_1^* y/c))^m\, f(y)\, dy, \tag{4.21}$$

where χ^2 is the distribution function of a chi-squared random variable with $2(n-p+1)$ degrees of freedom and $f(y)$ is the probability density function of a chi-squared random variable with $2p$ degrees of freedom.

5. TABLES AND AN EXAMPLE

Using MATHEMATICA, the minimum probability of a correct selection under the least favorable configuration $P(\mathrm{CS} \mid \mathrm{LFC})$ is computed. The least favorable configuration given in (4.16) is usually called the generalized least favorable configuration, since m, the number of the parameter λ smaller than δ_1, is not known. For each value of k, there are $k+1$ possible values for m. Thus there are $k+1$ possibilities for the LFC. For p (the number of components in a signal) $=5$, 10, and 20; $\delta_1^*/c = 1/2$, 1/3, 1/4, and 1/5; $\delta_2^*/d = 2$, 3, 4, and 5; $P^* = 0.90$, and 0.95, we computed the integral in Eq. (4.21) for $m = 0, 1, \ldots, k$ for $k = 4$. The purpose of this computation is to find the minimum sample sizes such that the P^* requirement is met. For each case considered, the smallest value of $P(\mathrm{CS} \mid \mathrm{LFC})$ over the $k+1$ values of m is always (without any exception in the cases we considered) a unimodal function in n.

To illustrate this property and the method used to produced the tables, we present in Figs. 1 through 5 for the cases $m = 0, 1, \ldots, 4$, respectively, the plots of the integral given in Eq. (4.21) as a function of sample size n for the special configuration $p = 10$, $\delta_1^*/c = 1/3$, and $\delta_2^*/d = 3$. From these figures, it is clear that the $P(\mathrm{CS} \mid \mathrm{LFC})$ is always unimodal. This was true for all cases considered. Furthermore, presented in Fig. 6 are minimum values of the five integrals (for $m = 0$ to 4) versus sample size n. These results are for the same cases considered in Figs. 1 to 5. In Fig. 6, $y = 0.90$ and $y = 0.95$ provide the sample size n corresponding to desired values of P^*. From these figures it is clear that the minimum of the $k+1$ integrals is also unimodal in n. Thus the line $y = 0.90$ intersects the curve representing the minimum of $k+1$ integrals at either 0 or, at most, two points. The first intersection provides a solution for the sample size requirement (n).

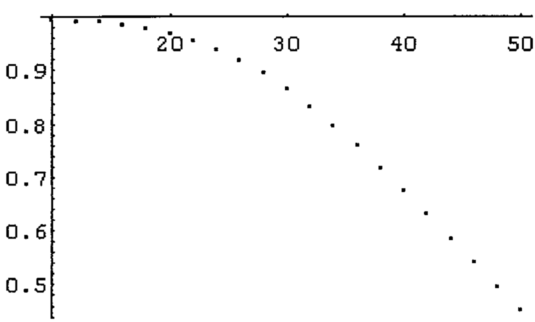

FIGURE 1

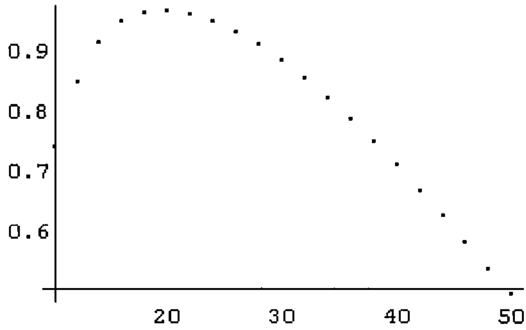

FIGURE 2

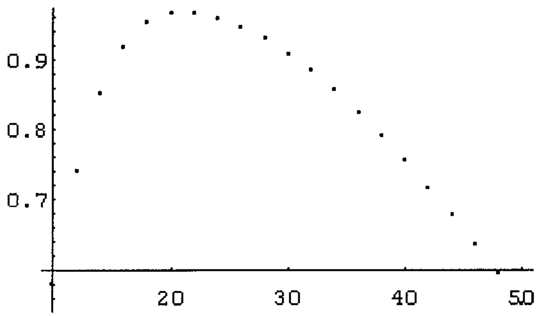

FIGURE 3

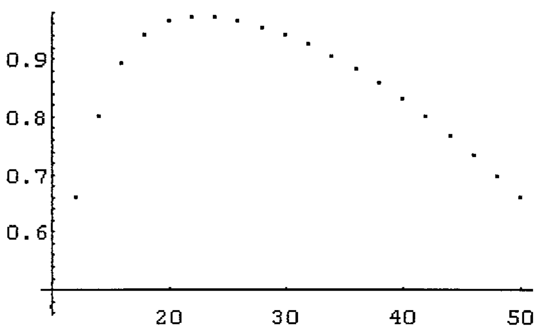

FIGURE 4

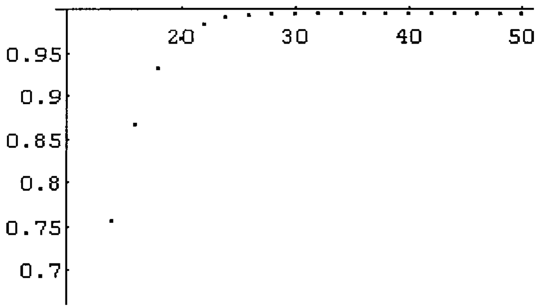

FIGURE 5

Table 1 presents the sample sizes required to satisfy P^* requirements for $k=4$. We found in our reserach that the minimum of the $k+1$ integrals will not always occur at predictable values of m. Furthermore, the minimum may not be bounded by P^* for all values of m. In this situation, adjustments to the parameters δ_1^*/c and δ_2^*/d may be required. Other statistical techniques may be required such as changes to the value dimension p, in order to satisfy the requirement on P^*. In the table, "*" is used to indicate that there is no solution for the given sample size n. In general, as p increases, we should always be able to find a solution.

This next example demonstrates use of the table for analysis of radar data. To illustrate the proposed selection procedure, five 20-variate complex multivariate normal populations are generated using a MATLAB-based radar simulation tool, developed for the United States Air Force Rome Laboratory by Scientific Studies Corporation. Fifty test data, $\mathbf{x}$, are generated from population π_0 with covariance matrix Σ_0. Samples of size 40 are generated from four other populations π_1, π_2, π_3, and π_4 with covariance matrices Σ_1, Σ_2, Σ_3, and Σ_4. The distance function for these

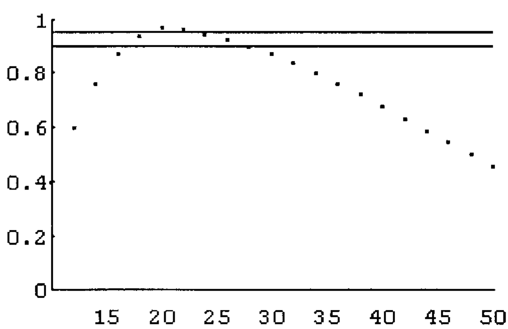

FIGURE 6

TABLE 1

The Sample Size n Needed to Achieve the P^* Requirement for $k=4$

	$P^*=0.90$						$P^*=0.95$			
$\delta_1^*/c \backslash \delta_2^*/d$	2	3	4	5		$\delta_1^*/c \backslash \delta_2^*/d$	2	3	4	5
					$p=5$					
$\frac{1}{2}$	*	*	*	12		$\frac{1}{2}$	*	*	*	*
$\frac{1}{3}$	*	*	10	10		$\frac{1}{3}$	*	*	*	*
$\frac{1}{4}$	*	*	9	9		$\frac{1}{4}$	*	*	*	10
$\frac{1}{5}$	*	*	9	9		$\frac{1}{5}$	*	*	*	9
					$p=10$					
$\frac{1}{2}$	*	21	21	21		$\frac{1}{2}$	*	*	22	22
$\frac{1}{3}$	*	18	18	18		$\frac{1}{3}$	*	19	19	19
$\frac{1}{4}$	16	16	16	16		$\frac{1}{4}$	*	17	17	17
$\frac{1}{5}$	15	15	15	15		$\frac{1}{5}$	*	16	16	16
					$p=20$					
$\frac{1}{2}$	38	38	38	38		$\frac{1}{2}$	40	40	40	40
$\frac{1}{3}$	33	33	33	33		$\frac{1}{3}$	34	34	34	34
$\frac{1}{4}$	30	30	30	30		$\frac{1}{4}$	32	32	32	32
$\frac{1}{5}$	29	29	29	29		$\frac{1}{5}$	30	30	30	30

Note. A "*" sign shows that the probability requirement P^* is not satisfied by any sample size.

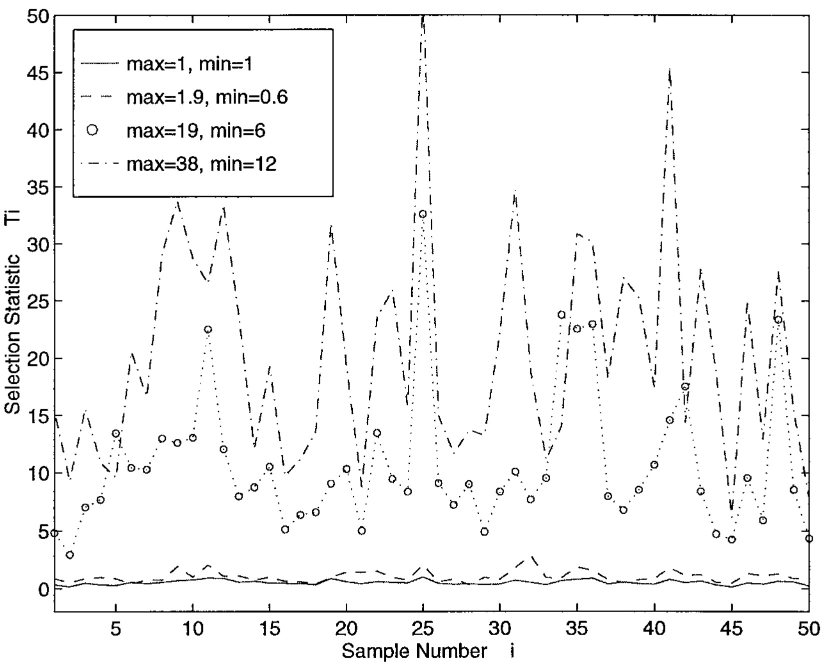

FIG. 7. Figure for example.

covariance matrices satisfies the requirement that the largest and the smallest eigenvalues of $\Sigma_0 \Sigma_j^{-1}$ for $j=1$, 2, 3, and 4 are (1.0, 1.0), (1.9, .6), (19, 6), (38, 12) respectively. Using results from Table 1, for the case $k=4$, $p=20$, $P^*=0.95$, $\delta_1^*=1/6$, and $\delta_2^*=6$, procedure parameters $c=0.5$ and $d=2$ are required. The selection statistic $T_i = x^H S_i^{-1} x/n$ for each of the 50 test data is presented in Fig. 7. It is clear from the figure that the populations π_1 and π_2 are always selected as similar (homogenous) populations; population π_3 and population π_4, the nonhomogenous populations, and never selected. Notice that a "correct screening" is defined as the rejection of a dissimilar population. The screening of a good population is not necessarily incorrect as long as only good populations remain after applying the screening procedure. The proportion of correct screening is 100% for the 50 test data analyzed in this example.

6. CONCLUDING REMARKS AND DIRECTIONS FOR FUTURE RESEARCH

Limitations on sample size often prohibits accurate estimation for the covariance structure. Since the statistic $T_i = x^H S_i^{-1} x/n$ provides information about the deviation of the covariance matrix Σ_i from Σ_0, it is used to develop a screening procedure for the analysis of multivariate data. The underlying distribution for both the test (primary) data and the reference (secondary) is assumed to have zero mean. Since the sample covariance matrix is not defined for a single observation when the mean vector is unknown, the selection statistic $T_i = x^H S_i^{-1} x/n$ provides a suitable alternative.

Results presented in this paper demonstrate that the statistic $T_i = x^H S_i^{-1} x/n$ may be used to identify secondary data which have covariance matrices different from those for a control population. The principal results obtained in this paper are (1) a new selection formulation for screening secondary data in radar signal processing; (2) an inferential selection procedure to achieve the goal of screening the data; and (3) formulas for computing the performance measure, $P(\mathrm{CS})$ for the proposed procedure. A numerical example illustrates this procedure.

In order to further assess performance of this selection procedure for radar applications, measures of robustness are required. From the theoretical developments in Section 4, it is possible to find procedure parameters such that, with high probability P^*, a correct selection or screening is always obtained. If a statistical hypothesis test, designed under the assumption that only homogenous secondary data are available for parameter estimation, is applied using screened data which have been identified via this new selection procedure, then it is reasonable to assume that detection performance will improve. Whether this improvement in radar signal processing performance is significant enough to warrant further development is an important issue. This issue is being addressed, and will form the basis for future publications.

ACKNOWLEDGMENTS

The research was conducted while P. Chen was a visiting scientist at the US Air Force Rome Laboratory. The authors thank the Air Force Office of Scientific Research for support throughout this effort.

REFERENCES

T. W. Anderson, "An Introduction to Multivariate Statistical Analysis," 2nd ed., Wiley, New York, 1984.

R. E. Bechhofer, A single-sample multiple decision procedure for ranking means of normal populations with known variances, *Ann. Math. Statist.* **25** (1954), 16–39.

S. S. Gupta, On a decision rule for a problem of ranking means, Mimeograph Series No. 150 Institute of Statistics, University of North Carolina, Chapel Hill, NC, 1956.

S. S. Gupta and S. Panchapakesan, Some selection and ranking procedures for multivariate normal populations, *in* "Multivariate Analysis, II" (P. R. Krishnaiah, Ed.), pp. 475–505, Academic Press, New York, 1969.

S. S. Gupta and S. Panchapakesan, "Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations," Wiley, New York, 1979.

H. Hotelling, A generalized T test and measure of multivariate dispersion, *in* "Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability" (J. Neyman, Ed.), pp. 23–41, Univ. of California Press, Los Angeles/Berkeley, 1951.

A. T. James, Distributions of matrix variates and latent roots derived from normal samples, *Ann. Math. Statist.* **35** (1964), 475–501.

E. J. Kelly, An adaptive detection algorithm, *IEEE Trans. Aerospace Electron. Systems* **22**, No. 1 (1986), 115–127.

C. G. Khatri, Some distribution problems connected with the characteristic roots of $\boldsymbol{S}_1\boldsymbol{S}_2^{-1}$, *Ann. Math. Statist.* **38** (1967), 944–948.

C. G. Khatri and C. R. Rao, Test for a specified signal when the noise covariance matrix is unknown, *J. Multivariate Anal.* **22**, No. 2 (1987), 177–188.

W. L. Melvin, M. C. Wicks, and P. Chen, Nonhomogeneity Detection Method and System for Improved Adaptive Signal Processing, U.S. patent 5706013, 1998.

R. S. Raghavan, H. E. Qiu, and D. J. McLaughlin, CFAR detection in clutter with unknown correlation properties, *IEEE Trans. Aerospace Electron. Systems* **31**, No. 2 (1995), 647–657.

C. R. Rao, "Linear Statistical Inference and Its Applications," 2nd ed., Wiley, New York, 1973.

J. R. Roman and D. W. Davis, "Multichannel System Identification and Detection using Output Data Techniques," Final Report No. SSC-TR-96-02, Vol. II, Scientific Studies Corporation, 1996

S. N. Roy, The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null hypotheses, *Sankhya* **7** (1945), 133–158.

M. S. Srivastava and C. G. Khatri, "An Introduction to Multivariate Statistics," Elsevier/North-Holland, New York, 1979.